ASTROBIOLOGY
Volume 7, Number 6, 2007

DOI: 10.1089/ast.2006.0091

Hypothesis Paper

Did Earthquakes Keep the Early Crust Habitable?

NORMAN H. SLEEP and MARK D. ZOBACK

ABSTRACT

The shallow habitable region of cratonal crust deforms with a strain rate on the order of $\sim 10^{-19}$ s^{-1}. This is rapid enough that small seismic events are expected on one-kilometer spatial scales and one-million-year timescales. Rock faulting has the potential to release batches of biological substrate, such as dissolved H_2, permitting transient blooms. In addition, the steady-state deformation of the brittle crust causes numerous small faults to be permeable enough (on the order of $\sim 10^{-15}$ m^2) for water to flow on a kilometer scale over relatively short geological times ($\sim 10^5$ yr). Hence, active faults act as concentrated niches capable of episodically tapping resources in the bulk volume of the rock. Radiolysis and ferrous iron are potentially bases of sustainable hard-rock niches. Key Words: Methanogens—Extreme environments—Cratons—Continental tectonics. Astrobiology 7, 1023–1032.

INTRODUCTION

Terrestrial cratons are regions of relative geological inactivity and provide an analogue to one-plate planets such as Mars. The crust of these regions is inhabited down to the depths of the deepest mines, currently 4.1 km. Microbes at these depths obtain their energy from the rock. They do not depend on the products of photosynthesis at the surface.

The purpose of this paper is to follow a suggestion by Sherwood Lollar *et al.* (2006) that episodic fracturing in the crust leads to the episodic release of H_2- and CH_4-rich abiogenic gas. This process parcels out an energy resource to biota. We use geodynamics to quantify processes that maintain habitability over long geological periods. We begin with the thermodynamic requirements for life. We then examine stress and faulting within the crust and its effect on rationing and gathering resources. Finally, we address flow and sequestration of hydrous fluids, which is a major part of the process.

THERMODYNAMICS AND LIFE

Life needs to gather chemicals and energy from its environment. It can assimilate complex organic compounds already present, as we do when we obtain vitamin C from lemons. It can capture photons, produce complex organic compounds, and store energy for later use, as does the lemon tree. Life can also use chemical disequilibria present in its environment as an energy source. Only the last option is available to microbes in deep hard rocks.

Department of Geophysics, Stanford University, Stanford, California.

Gibbs energy requirements

Following Hoehler *et al.* (1998, 2001) and Hoehler (2005), we use the methanogen reaction as a generic example of a deep niche. Hydrogen and carbon dioxide react to form methane and water

$$CO_2 + 4H_2 \Rightarrow CH_4 + 2H_2O. \quad (1)$$

This reaction provides Gibbs energy, G. To benefit, the methanogens must couple with a Gibbs-energy consuming reaction, such as ADP (Adenosine diphosphate) ⇒ ATP (Adenosine triphosphate).

The need to couple an energy-consuming reaction limits habitability. To illustrate this effect, we normalize Reaction 1 to one mole of H_2 and assume that the reactants are in aqueous solution (*i.e.*, $0.25CO_2 + H_2 = 0.25CH_4 + 0.5H_2O$). We vary only the activity of H_2 in Reaction 1 and express the activity of hydrogen in equilibrium with the biological reaction:

$$A_{H2} = (A_{H2})_A \exp(\Delta G_B/RT) \quad (2)$$

where $(A_{H2})_A$ is the partial pressure in equilibrium with the uncoupled reaction when all other reactants are held constant, ΔG_B is the extra Gibbs energy needed for the coupled reaction, R is the gas constant, and T is absolute temperature.

The reaction of ADP ⇒ ATP requires ~36 kJ mol^{-1} (*e.g.*, Hoehler *et al.*, 2001). The Gibbs energy associated with methanogenesis appears to be small enough compared with this so that (2) poses real limits to habitability. Microbes draw down H_2 to the thermodynamic limit if the reactant CO_2 is present in excess. The concentration of the remaining substrate and the amount of remaining Gibbs energy are variable and depend on species and substrate (*e.g.*, Jackson and McInerney, 2002). For methanogens, Hoehler *et al.* (2001) report 10.6 kJ mol^{-1} per methane in Reaction 1 for wild archaea, which is 2.65 kJ mol^{-1} of hydrogen. In contrast, methanogens investigated in the laboratory by Kral *et al.* (1998) require ~3 times as much energy (~33 kJ mol^{-1} per methane). The coupled reactions use ~12 (Hoehler *et al.*, 2001) and ~4 (Kral *et al.*, 1998) H_2 per ATP, respectively. The final H_2 concentration is small but not zero. For example, hot spring methanogens investigated by Chapelle *et al.* (2002) deplete hydrogen to ~13 nmol. Our conclusions do not depend on the precise value of this amount.

Consortia of organisms, where the product of one is the substrate for the next, lower the concentration of final substrate and remaining Gibbs energy (Jackson and McInerney, 2002). This option is unavailable to methanogens in rocks, as their product methane is the minimum Gibbs energy compound that cannot serve as a substrate in the absence of an oxidant.

Rationing and sustainability

The total amount of energy available from a block of crustal rock is limited. For the crust to be habitable over geological time, the supply of energy to microbes cannot be too fast or too slow. We begin with the former issue. Consider that a homogeneous region has water with abundant H_2 and CO_2. If there are also plenty of nutrients, the methanogen population blooms and consumes the reactants in (1) until the limiting reactant is depleted to the thermodynamic limit in (2). If no more reactants arrive, the methanogens eventually die off. An organism might live on their remains, but the region would become sterile over geological time.

Neither is a very slow supply of the reactants good for microbes. Abiotic reactions may deplete a reactant below the thermodynamic limit in (2). Even if this does not occur, the flux of reactants may be too low to sustain the energy requirement of the microbes. Finally, the microbes occasionally need to have a surfeit of energy to reproduce. Otherwise, they cannot evolve. They will die off individually when local conditions become untenable. Chemical, spatial, and temporal mechanisms that naturally ration and episodically dole out reactants keep microbes from this fate.

Biota in ancient deep-sea sediments illustrate natural rationing mechanisms (*e.g.*, Parkes *et al.*, 2000). Slow ingress of oxidants limits consumption in organic-rich sediments. Microbes very slowly consume refractory organic matter where oxidants are available.

Interfaces provide local lush environments for subsurface biota that do not rapidly deplete large reservoirs (*e.g.*, Parkes *et al.*, 2000). In the case of fractures in rocks, the microbes may live with a significant number of cells per volume of water but consume resources at a slow rate per volume of bulk rock. A zone of mixing between two fluids in a fracture system is a small, but potentially bountiful, niche.

Subsurface energy sources

In the case of deep biota discussed by Sherwood Lollar *et al.* (2006), sealed cracks serve to sequester H_2-rich water, and their rupture leads to occasional bounty. Some of the hydrogen comes from radiolysis from natural radioactivity in the rock (Lin *et al.*, 2005). The resource is renewable over geological time.

All water-rock reactions are potentially exhaustible. For example, hydrogen is a product of the reaction of ferrous iron with water to make magnetite

$$3\text{FeO (in silicate or hydroxide)} + H_2O \Rightarrow Fe_3O_4\text{ (magnetite)} + H_2\text{ (aqueous)}. \quad (3)$$

Iron sulfide is another hydrogen source

$$\text{FeS (in rock)} + H_2S\text{ (aqueous)} \Rightarrow FeS_2\text{ (pyrite)} + H_2\text{ (aqueous)} \quad (4)$$

(Wachtershauser, 1988; Drobner *et al.*, 1990; Rickard, 1997; Rickard and Luther, 1997; Hoehler, 2005). These reactions are both potential energy sources for microbes and sources of hydrogen for methanogens. We do not consider more oxidized rocks, such as red granite, where the mineral buffer involves oxidation of magnetite to hematite. We also consider only iron-rich mafic and ultramafic rocks, which in aggregate are a significant fraction of cratonal crust and the crust of Mars.

We quantify exhaustibility by comparing the productivity from radiolysis to that from ferrous iron in rocks. The flux of H_2 per surface area is 8×10^{-6} mol m^{-2} yr^{-1} in the Witwatersrand Basin of South Africa (Lin *et al.*, 2005). There is bountiful energy in hard rock if water circulation brings in reactants and removes products. For example, there are 1.4×10^6 mol m^{-2} of H_2 potentially available from Reaction 3 from a 1 km column of mafic (or ultramafic) rock with 10% FeO by mass and density of 3000 kg m^{-3}. This quantity is large compared to current average-crust radiolysis flux. For example, it would provide a flux of 300×10^{-6} mol m^{-2} yr^{-1} if evenly distributed over the age of Earth.

The finite rate of water flow limits the rate of consumption of ferrous iron in basalt. For example, the H_2 concentration in freshwater Icelandic systems extrapolates to 10 μM at 100°C (Stefánsson and Arnórsson, 2002). As 1 km^3 is 10^{12} L, this concentration results in a volume water/rock ratio of 1.4×10^5 to remove the potentially available H_2 from the basalt. The end-member H_2 concentration from serpentinite in the marine Lost City vent is 15 m*M* (Kelley *et al.*, 2005), which implies a water/rock ratio of 10^2.

DYNAMICS OF STABLE CRUST

We now consider geodynamics to quantify processes that form fractures and new surface area, and the rate that water flows through them. We begin with the long-term deformation rate. We then obtain the rates of seismicity and fracturing, and the rate of groundwater flow through them.

Renewable and nonrenewable stresses and geological evidence of deformation

We can constrain the current rate of deformation of plate interiors from a variety of lines of evidence. From the perspective of geodesy, precise technology has been available for only a few years ($\sim 10^8$ s), and the breadth of plate interiors limits base lines to $\sim 10^4$ km. As modern geodetic techniques allow for the detection of baseline changes of 1 cm, we can resolve strain rates of 10^{-17} s^{-1}. Such platewide deformation is not observed in intraplate areas around the world (*e.g.*, Calais *et al.*, 2006). Transient deformation associated with glacial rebound is of this order and limits what might be learned about long-term deformation from more precise studies.

The geological stability of plate interiors provides one upper-bound constraint on intraplate deformation rates. From a geological perspective, if a strain rate of 10^{-17} s^{-1} were to persist for 1 billion years (3×10^{16} s), it would thin (or thicken) 40 km–thick crust by 12 km, which is unacceptable in a stable region, as such variations are not observed (Mooney *et al.*, 1998). It would also deform continents sufficiently over geological time so that the jigsaw of continents of plate tectonics would no longer fit together. A strain rate of 10^{-18} s^{-1} would cause a 1.2 km thickness change, which would still cause observable structures in some places. A strain rate less than this would not have obvious geological effects, given the stability of cratons. Hence, an upper limit on intraplate strain rates is 10^{-18} (Fig. 1).

We can also obtain a lower limit on strain rate by considering renewable and nonrenewable sources of intraplate stress. Body forces cause renewable stress. In analogy, consider a weight hanging from a viscoelastic spring. The spring creeps slowly over time but the weight maintains a constant force and constant stress on the spring. Now, consider the same spring stretched between 2 fixed attachments. The spring relaxes, leaving it unstressed.

Body forces, mainly from the buoyancy of mid-oceanic ridges (ridge push), and intraplate variations of lithosphere thickness and density (lithospheric buoyancy) maintain renewable stresses on the plates. The stress orientation from borehole measurements and earthquake mechanisms indicates that these forces dominate within the stable parts of the plate interiors (Zoback *et al.*, 1989). The strain rate from this stress must therefore exceed the strain rates for nonrenewable stress. As with a spring stretched between 2 attachments, kinematics constrains the strains and strain rate associated with nonrenewable stress. These processes include deformation from the change of the Earth's radius as the interior cools, changes in the ellipticity from tidal despinning, and north-south plate movement over the elliptic surface of Earth (Solomon, 1987). Combined, they lead to strain rates of a few 10^{-21} s^{-1}.

Another way to view this is in terms of the changes in the local geotherm over time, which generate thermoelastic stresses. Xenolith geotherm data provide some constraint on the rate that temperature changes (Bell *et al.*, 2003). Part of the change comes from the gradual cooling of Earth's sublithospheric mantle at ~50 K per billion years (Abbott *et al.*, 1994; Galer and Mezger, 1998). Intracontinental basin formation (*e.g.*, Kaminski and Jaupart, 2000) and mantle plumes may cause more rapid nonmonotonic changes. The strain rate scales to the strain at the depth where the material behaves rigidly, crudely 500°C, that is,

$$\varepsilon' = \frac{\lambda T_R \Delta T}{T_L \Delta t}, \quad (5)$$

where λ is the linear thermal expansion coefficient [$\sim 0.8 \times 10^{-5}$ K^{-1} for granite and $\sim 0.5 \times$

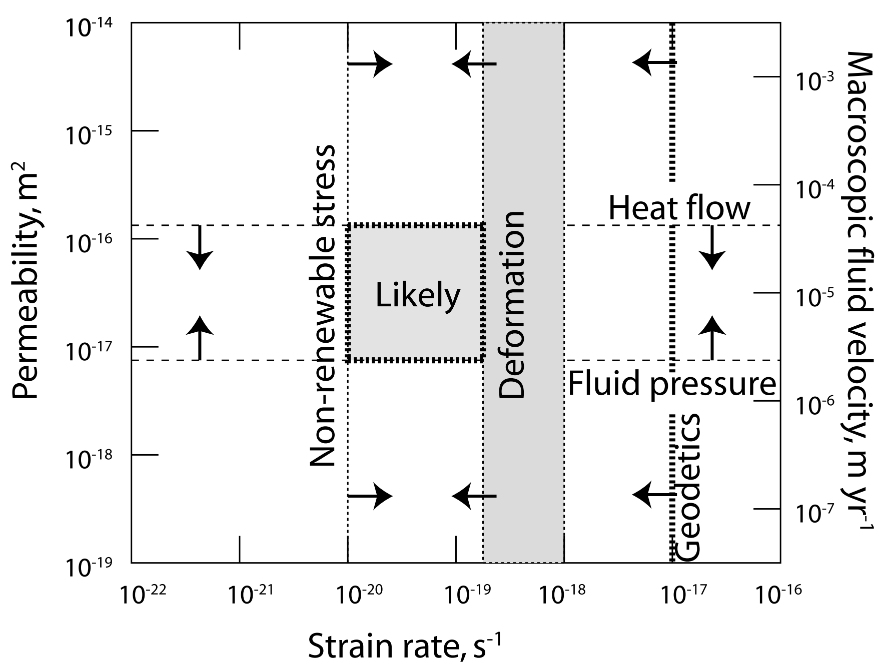

FIG. 1. Schematic diagram shows observational constraints on permeability and strain rate. The strain rate in cratons needs to be greater than 10^{-20} s^{-1}; otherwise nonrenewal stresses would dominate. It needs to be less than 10^{-18} because deformation of the crust is not evident. The permeability is greater than 10^{-17} m^2 because the fluid pressure is near hydrostatic. It needs to be less than 10^{-16} m^2 so that convection does not obviously perturb the geotherm from a conductive gradient in a deep well. These ranges imply sluggish but finite fluid flow favorable to sustained habitability.

10^{-5} K^{-1} for mafic rocks in the range 20–100°C (Skinner, 1966, Table 6-10)], T_R is the temperature where the lithosphere behaves rigidly, T_L is the temperature at the base of the lithosphere, and the temperature at the base of the lithosphere changes ΔT over time Δt. A change in the temperature at the base of the lithosphere of 50 K in a billion years causes a strain rate of 0.6×10^{-20} s^{-1}. Mantle plumes (*e.g.*, Sleep, 2006) would cause episodic strain rates a few times this amount.

Reiterating, the strain rate in cratons needs to be greater than strain rates that cause nonrenewable stress, that is, a few times, 10^{-20} s^{-1}. It needs to be less than ~10^{-18} s^{-1}, which would produce evident deformation over geological time (Fig. 1). As we deal only with processes that scale linearly with strain rate, we use a single value 10^{-19} s^{-1} in examples with the caveat that the actual strain rate is likely to vary in space and time. Zoback and Townend (2001) obtained 10^{-20} s^{-1} as an upper bound strain rate for stable regions, using rheological models for creep within the lithosphere and estimates of the magnitude of buoyancy-related forces.

The observed strains on Mars are of the same order as those in stable cratons on Earth, again with the caveat that observed strain rates vary in space and time. For example, the strain in Lunae Planum is 0.003 (Plescia, 1991), and the strain in Arcadia Planitia is 0.0006 (Plescia, 1993). Averaging these strains over 3 billion years gives rates of 0.6 to 3×10^{-20} s^{-1}. This is not surprising, as the processes are similar to terrestrial ones: global cooling of the interior, local cooling of the deep lithosphere, and body forces from lateral heterogeneities.

Stresses and strains within the habitable crust

Our nominal strain rate of 10^{-18} s^{-1} would produce significant stresses over geological time if it were not relieved. We consider the stress balance within the lithosphere to show what happens within the upper few habitable kilometers where temperatures are less than ~100°C.

Numerous boreholes have penetrated the habitable region of Earth's crust. Stress measurements, including induced seismicity, show that the crust is critically stressed so that it is near frictional failure at depths down to at least 8 kilometers (see summary in Townend and Zoback, 2000). That is, the resolved shear tractions on all the fractures satisfy the Coulomb inequality

$$\tau < \mu[S_N - P_F], \tag{6}$$

where μ is the coefficient of friction, 0.6–1.0. The term in the brackets is the effective stress, where S_N is the normal traction resolved on a fault surface and P_F is the fluid pressure. The shear traction is close to failure on many pre-existing faults in the brittle crust, even in stable continental interiors (Zoback *et al.*, 2002).

The Coulomb formula (6) suffices to show that stresses have relaxed in the upper crust. The normal traction in (6) scales with lithostatic pressure

$$S_L = \rho_r g Z, \tag{7}$$

where ρ_r is the density of the rock, g is the acceleration of gravity, and Z is depth. The hydrostatic pressure gradient is

$$P_H = \rho_w g Z, \tag{8}$$

where ρ_w is the density of water. Measurements indicate that the fluid pressure within the crystalline basement rocks of stable crust is very close to hydrostatic (see review in Townend and Zoback, 2000). Letting the acceleration of gravity equal 9.8 m s^{-2}, the coefficient of friction $\mu = 0.7$, the water density equal 1000 kg m^{-3}, and the rock density equal 2800 kg m^{-3} yields a shear stress in (6) of 62 MPa at 5 km depth. Deep brines in cratonal rocks are often saline (*e.g.*, Bottomley *et al.*, 2002, 2003; Fehn and Snyder, 2005; Négrel and Casanova, 2005). For example, the water density of 1100 kg m^{-3} implies a shear stress in (6) of 58 MPa at 5 km depth.

The ridge push buoyancy force imposes a stress resultant on the plate that determines the integral of the membrane stress over depth

$$Y = \int (S_{xx} - S_L) dz, \tag{9}$$

the indices xx indicate a component of the horizontal stress. The difference between horizontal and vertical stress causes creep in ductile materials and faulting if the Coulomb criterion (6) is exceeded. Faulting relaxes the stresses within the shallow part of the crust so that the stress difference in (9) satisfies (6). Creep relaxes stresses in deeper and hotter parts of the plate, typically ≥15 km depth. The stress in (9) thus concentrates in the middle part of the plate, where it is too deep for faulting and too cold for rapid creep. Ductile creep within this highly viscous region is the rate-limiting step. At steady state, the plate deforms over its full depth at a constant horizontal strain rate. This deformation generates little stress in the deep hot part of the plate and maintains the stress near frictional failure in the upper crust (Zoback *et al.*, 2002).

Earthquake recurrence

As noted above, earthquakes on faults maintain permeability in the crust and, in fact, episodically release trapped water by opening new pathways. The statistics of large earthquakes must be considered because earthquakes release much of the strain within the brittle zone, though we are mainly interested in small frequent events. For purposes of illustration, we bin earthquakes by 1 unit of magnitude as this leads to simple statistics. The Guttenberg and Richter relationship (*e.g.*, Kanamori, 1977; Marsan, 2005) implies that, for every 1 magnitude-7 event, there will be about 10 magnitude-6 events, 100 magnitude-5 events, and so on.

We use seismic moment to relate seismicity to the average cratonal strain rate of 10^{-19} s^{-1}. By definition, the moment is

$$M = \Gamma S A, \tag{10}$$

where Γ is the shear modulus (~40 GPa for crustal rocks), S is the displacement during the earthquake, and A is the surface area of fault rupture (Kanamori, 1997). From Hooke's law, the slip on the fault is proportional to its dimension $x \equiv \sqrt{A}$ and the stress drop $\Delta\sigma$. This gives the moment in terms of stress drop, which is empirically independent of earthquake size

$$M = x^3 \Delta\sigma. \tag{11}$$

The strain averaged over some crustal volume V is

$$\varepsilon = \frac{M}{\Gamma V} = \frac{SA}{V}. \tag{12}$$

The moment magnitude m is defined by

$$M = (1.25 \times 10^9 \text{ nt-m}) m^{1.5}. \tag{13}$$

(Kanamori, 1977). For example, (11) and (13) imply that a magnitude-2 event with a typical stress drop of 1.25 MPa has a surface area of 10^4 m^2. Because there is potentially a great deal of rock fracturing and comminution accompanying earthquakes, the surface area of fault slip is a lower bound on the surface area created by a slip event.

We are now ready to use the globally observed relationship that the number of events in a magnitude bin decreases by a factor of 10 for each magnitude, m. The total moment release is

$$M_T = C_m \int_{-\infty}^{m_{\max}} 10^{0.5\mathrm{m}}\, dm, \tag{14}$$

where C_m is a constant, $m_{\max}$ is the magnitude of the largest event, and the integral converges rapidly enough that the lower limit is taken as $-\infty$. The moment released by events with magnitudes between $m_{\max}$ and $m_{\max} - \Delta m$ is

$$M(\Delta m) = C_m \int_{m_{\max}-\Delta m}^{m_{\max}} 10^{0.5\mathrm{m}}\, dm = M_T\,[1 - 10^{-0.5m}]. \tag{15}$$

(*e.g.*, Marsan, 2005). This implies that the largest events in a region release most of the moment. Conversely, the total moment release is statistically the moments of the earthquakes in the maximum moment bin ($\Delta m = 1$) divided by $1 - \sqrt{10}$. In addition, the Guttenberg and Richter relationship between strain and moment release (12), and the relationship between moment and magnitude (13) imply that total surface area (the area A of an individual event times the number of events with that magnitude) is the same in each magnitude bin.

The number of small events, however, is not strongly dependent on the magnitude of the largest events. We let the largest magnitude of intraplate earthquakes in most regions be either 6 or 7. The number of magnitude-2 events in a cubic kilometer per million years is 6 and 2, respectively, for a strain rate of 10^{-18} s^{-1}. There are between 600–200 magnitude-0 events, each with a surface area of 100 m^2 and 60,000–20,000 magnitude-2 events each with a surface area of 1 m^2. The total ruptured area depends linearly on the number of magnitude bins that actually occur. We assume 5 bins with magnitude 2 being the largest that can be reasonably expected on a kilometer scale over a reasonably short geological time and magnitude 2 as a cut off for small events. This gives a lower bound of surface-area production rate of 600,000–200,000 m^2 per million years.

Biota directly benefiting from earthquake rupture (releasing, for example, H_2) thus have to persist over times between events. This conclusion holds even though our recurrence times are somewhat pessimistic. The actual distribution of events in space is patchy. Faults favorably oriented for slip in a given stress field would be expected to rupture repeatedly in otherwise intact rock.

Permeability and fluid flow

We now consider what happens during the interseismic period in continental crust. Direct measurements indicate that the crust is permeable enough that pore water is essentially at hydrostatic pressure and fractures aligned favorably to the stress field carry most of the fluid flow (Townend and Zoback, 2000; Zoback and Townend, 2001).

Circulating flow is relevant to biota, as it brings in reactants such as CO_2 and sweeps out reaction products such as CH_4. As explained in the Appendix, it would be inappropriate to consider the stability of convection in a permeable laterally homogeneous region. We show here that heterogeneity implies that a small amount of hydrothermal circulation always occurs.

We begin with D'Arcy's law; the macroscopic velocity of flow is dimensionally

$$V_D = \frac{k \Delta \rho_w g}{\eta_w}, \tag{16}$$

where k is the permeability, $\Delta\rho_w$ is the local lateral variation of water density from lateral variations in temperature, and η_w is the viscosity of the water. The particle velocity of the water is

$$V_P = \frac{V_D}{\phi}, \tag{17}$$

where ϕ is the through-going porosity.

We give an example for the 5 km–deep base of the habitable zone where the temperature is $\sim$100°C and the fluid pressure is 50 MPa. The viscosity is 0.3×10^{-3} Pa s. Our treatment also applies to Mars if water is present, as we obtain order-of-magnitude answers. Lateral variations in temperature result from lateral variations in thermal conductivity, lateral variations in radioactive element distribution, and lateral variations in the heat flow coming from below. Lateral variation of a few K over the distance of a kilometer is reasonable at this depth. To provide a quantitative example, we assume lateral temperature variations of 5 K (Clauser *et al.*, 1997; Popov *et al.*, 1999), which implies density variations 3 kg m^{-3} for the volume thermal expansion coefficient (6×10^{-4} K^{-1}) of water at hydrostatic pressure at 5 km depth and 100°C. The permeability is between 10^{-16} and 10^{-17} m^2 (Townend and Zoback, 2000). This implies a macroscopic velocity of 10^{-11} and 10^{-12} m s^{-1} or 3×10^{-4} to 3×10^{-5} m yr^{-1}, respectively (Fig. 1). The fluid velocity within a given fracture is a factor of the inverse of the open crack porosity times the macroscopic velocity, 100 to 1000 times greater. This gives a particle velocity range of 3–300 m per millennium. This implies that the water is not stagnant and that substrates, such as H_2, which are produced at depth and are soluble in water, readily move around on a short geological timescale once they are in major cracks. Hence faulting opens new pathways, which allow circulating fluids to "mine" the nutrients. The bounty is likely to cause microbial blooms. Unfaulted domains in the crust have permeability on the order of 10^{-19} to 10^{-20} m^2 (Townend and Zoback, 2000). These rocks sequester nutrients but at such low permeabilities that fluid movement would be so slow as not to provide sufficient nutrients to sustain life. Conversely, if permeability were much higher than 10^{-16} m^2, fluid movement would be so rapid as to exhaust nutrients from a given volume of rock in short geological periods of time. We show in the Appendix that such high permeabilities can be ruled out by observations of conductive heat flow to great depths in the crust.

Studies of deep brines provide evidence of the properties of deeply trapped water. Iodine-129 studies indicate that brines have remained trapped (that is, not mixed with surface water) for tens of millions of years (Bottomley *et al.*, 2002; Fehn and Snyder, 2005). Although the origin of the brines is not clear, it is evident that grain-scale permeability was high enough that extensive water-rock reaction occurred (Bottomley *et al.*, 2002, 2003; Négrel and Casanova, 2005; Fehn and Snyder, 2005).

Resource sustainability

We now use these computed flow rates to appraise whether hard rocks have enough resources to sustain life for geologically long times. We start with the water to rock ratio. Dimensionally, the volume ratio is

$$R = \frac{V_D t}{Z}, \tag{18}$$

where the circulation persists for time t, and an equivalent thickness Z is available for reaction. For 1 billion years, the computed flow rates, and a thickness of 5 km, the volume ratio is 100–1000. The ratio for stagnant brines is lower. For example, 1% porosity and a residence time of 10 million years (see Bottomley *et al.*, 2002; Fehn and Snyder, 2005) imply a ratio of 1 over a billion years.

We illustrate the implications of this ratio with an example for basalt. The reactant (10% by mass FeO) is present on the order of 1 mole per liter of rock. The product H_2 is released at 10 μM of water using the Icelandic data (Stefánsson and Arnórsson, 2002). Our higher estimated water/rock ratio, 1000, depletes only 1% of this resource.

Water flow, however, is able to deplete rocks that are more reactive than basalt, such as serpentinite. For example, the H_2 concentration of end-member Lost City marine vent fluid is 15 mM (Kelley *et al.*, 2005). As already noted, a water/rock ratio of 100 would suffice to remove the H_2 produced by the ~10% FeO in the serpentinite rock.

The crack heterogeneity of the rock tends to ration such potentially exhaustible resources. Crack margins react, but intact masses of rock stay fresh until tapped by new fractures. Trapped water equilibrates with nearby grains, as indicated by old brines (Bottomley *et al.*, 2002; Fehn and Snyder, 2005), but does not transport reactants.

CONCLUSION

We began with the suggestion by Sherwood Lollar *et al.* (2006) that episodic faulting releases batches of H_2-rich or CH_4-rich water into hard-rock environments. This transiently puts the reactant well above the level that abiotic processes can consume it. The microbial population blooms during the temporary period of bounty. The probable strain rate in stable crust is rapid enough that numerous cracks fail in a cubic kilometer volume on a million year timescale. The mechanism is attractive for maintaining habitability in the upper 5 km or so of the cratonic crust.

The water and the organisms within the deep crust should not be regarded as spatially stagnant. The stress in the habitable zone also maintains a population of permeable cracks, which are local niches that tap large volumes of bulk rock. The water moves on the scale of kilometers over millennium to million-year timescales. The cumulative water/rock ratio over geological time is large, 100–1000 by volume. There is enough FeO in basalt to provide hydrogen to biota over billions of years of geological time.

The situation for microbes in the crust would differ significantly if the permeability and flow rate were much larger or much smaller than we assumed. Lower flow rates imply a much smaller sustainable population. Higher flow rates imply thermal convection with hot upwelling regions and cool downwellings. The available resources, especially within upwellings, could be exhausted over relatively short periods of geological time.

APPENDIX: FREE CONVECTION

Rapid fluid circulation would significantly perturb the temperature field in the crust if it were significantly more permeable than we have assumed. We summarize the formalism with simple equations.

In the absence of hydrothermal circulation, conduction would carry heat from Earth's interior to its surface. To a first order, the vertical heat flow,

$$q = K \frac{\partial T}{\partial Z}, \tag{19}$$

where K is the thermal conductivity and Z is depth, would not vary much with depth. In fact, deep boreholes in hard rock show little variation of heat flow with depth. Deviations of the observed geothermal gradient in bores from the geothermal gradient implied by constant heat flow are modest, that is, several K (Clauser *et al.*, 1997; Popov *et al.*, 1999). This observation provides an upper limit on the rate of fluid flow and an upper limit on permeability.

The convective heat flow is dimensionally

$$q_v \approx \rho_w C_w V_D \Delta T \tag{20}$$

where C_w is the specific heat of the water ~4×10^3 J kg^{-1} K^{-1}, ρ_w is the density of water, and ΔT is the temperature contrast between the water and the surface. The Darcy velocity in terms of water properties is

$$V_D = \frac{k \alpha_w \rho_w g \Delta T}{\eta_w}, \tag{21}$$

where α_w is the volume thermal compaction coefficient of water. Combining (20) and (21) yields

$$q_v \approx \frac{\rho_w^2 C_w \alpha_w k g \Delta T^2}{\eta_w} \quad (22)$$

It is illustrative to rewrite (22) in terms of the dimensionless Raleigh number Ra

$$q_v \approx \left[\frac{k\Delta T}{Z}\right]\left[\frac{\rho_w^2 g C_w \alpha_w \Delta T k Z}{\eta_w K}\right] \equiv q_d \Delta T \quad (23)$$

where Z is the distance between 2 horizontal isothermal boundaries and ΔT is the temperature contrast between the boundaries. When the Raleigh number is below a critical value, convection does not occur within a homogeneous region. In Earth, local lateral temperature gradients associated with heterogeneities drive sluggish convection. Convection occurs within a homogeneous region when the Raleigh number exceeds the critical value. If the Raleigh number modestly exceeds the critical value, the heat flux by vigorous convection is proportional to the Raleigh number. The critical Raleigh number is $4\pi^2 = 39.5$ for 2 permeable boundaries and 27.1 for a lower impermeable boundary and a constant pressure upper boundary (Turcotte and Schubert, 2002).

In the case of the upper continental crust, the thermal state of Earth's interior determines the heat flow from below; it is $\sim$40 mW m^{-2} within stable shields (*e.g.*, Kaminski and Jaupart, 2000). The thermal gradient is then $\sim$20 K km^{-1} as we assumed above. The permeability at the critical Raleigh number in terms of a given conductive thermal gradient from (23) is

$$k_{\text{crit}} = \frac{4.2 \times 10^{-10}}{Z^2}\left(\frac{\partial T}{\partial Z}\right)^{-1} \quad (24)$$

where all the terms are in SI units and the term in brackets is the conductive thermal gradient (Turcotte and Schubert, 2002, Equation 9-133). The critical permeability depends strongly on the thickness of the convective region: 8.3×10^{-16} for a 5 km–thick layer and 0.93×10^{-16} for a 15 km layer. We can exclude a through-going permeability of more than a few 10^{-16} m^2 as all deep borehole data available confirm a nearly conducting geothermal gradient (Clauser *et al.*, 1997; Popov *et al.*, 1999).

Finally, the existence of dense brine in the deep subsurface (Bottomley *et al.*, 2002, 2003; Négrel and Casanova, 2005; Fehn and Snyder, 2005) provides stable stratification. Modest temperature differences, tens of Kelvin, cannot cause dense brines to ascend into less dense fresh waters. However, deep dense brines can thermally convect internally if they are not locally stratified. In this case, the brine may have circulated considerably in the subsurface yet be old in the sense that it has not mixed with surface reservoirs.

Thermohaline convection is complicated and beyond the scope of this paper. We note that episodes of rapid overturn can occur when the density change from temperature gradients overwhelms the chemical stratification. Conversely, local unstable differences in salinity may drive flow.

REFERENCES

Abbott, D.L., Burgess, L., Longhi, J., and Smith, W.H.F. (1994) An empirical thermal history of the Earth's upper mantle. *J. Geophys. Res.* 99, 13835–13850.

Bell, D.R., Schmitz, M.D., and Janney, P.E. (2003) Mesozoic thermal evolution of the southern African mantle lithosphere. *Lithos* 71, 273–287.

Bottomley, D.J., Renaud, R., Kotzer, T., and Clark, I.D. (2002) Iodine-129 constraints on residence times of deep marine brines in the Canadian Shield. *Geology* 30, 587–590.

Bottomley, D.J., Chan, L.H., Katz, A., Starinsky, A., and Clark, I.D. (2003) Lithium isotope geochemistry and origin of Canadian Shield brines. *Ground Water* 41, 847–856.

Calais, E., Han, J.Y., DeMets, C., and Nocquet, J.M. (2006) Deformation of the North American plate interior from a decade of continuous GPS measurements. *J. Geophys. Res.* 111, B06402.

Chapelle, F.H., O'Neill, K., Bradley, P.M., Methé, B.A., Ciufo, S.A., Knobel, L.L., and Lovley, D.R. (2002) A hydrogen-based subsurface microbial community dominated by methanogens. *Nature* 415, 312–315.

Clauser, C., Giese, P., Huenges, E., Kohl, T., Lehmann, H., Rybach, L., Šafanda, J., Wilhelm, H., Windloff, K., and Zoth, G. (1997) The thermal regime of the crystalline continental crust: implications from the KTB. *J. Geophys. Res.* 102, 18417–18441.

Drobner, E., Huber, H.J., Wachterhauser, G., Rose, D., and Stetter, K.O. (1990) Pyrite formation linked with hydrogen evolution under anaerobic conditions. *Nature* 346, 742–744.

Galer, S.J.G. and Mezger, K. (1998) Metamorphism denudation and sea level in the Archean and cooling of the Earth. *Precambrian Res.* 92, 387–412.

Fehn, U. and Snyder, G.T. (2005) Residence times and source ages of deep crustal fluids: interpretation of ^{129}I and ^{36}Cl results from the KTB-VB drill site, Germany. *Geofluids* 5, 42–51.

Hoehler, T.M. (2005) Biochemistry of dihydrogen (H_2). In

Metal Ions in Biological Systems, Volume 43: Biogeochemical Cycles of Elements, edited by A. Sigel, H. Sigel, and R.K.O. Sigel, Taylor and Francis, Boca Raton, pp. 9–48.

Hoehler, T.M., Alperin, M.J., Albert, D.B., and Martens, C.S. (1998) Thermodynamic control on H_2 concentrations in an anoxic marine sediment. *Geochim. Cosmochim. Acta* 62, 1745–1756.

Hoehler, T.M., Alperin, M.J., Albert, D.B., and Martens, C.S. (2001) Apparent minimum free energy requirements for methanogenic Archaea and sulfate-reducing bacteria in an anoxic marine sediment. *FEMS Microbiology Ecology* 38, 33–41.

Jackson, B.E. and McInerney, M.J. (2002) Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature* 415, 454–456.

Kaminski, E. and Jaupart, C. (2000) Lithospheric structure beneath the Phanerozoic intracratonal basins of North America. *Earth Planet. Sci. Lett.* 178, 139–149.

Kanamori, H. (1977) The energy release in great earthquakes. *J. Geophys. Res.* 82, 2981–2987.

Kelley, D.S., *et al.* (2005) A serpentine-hosted ecosystem: the Lost City hydrothermal field. *Science* 307, 1428–1434.

Kral, T.A., Brink, K.M., Miller S.L., and McCay, C.P. (1998) Hydrogen consumptions by methanogens on the early Earth. *Orig. Life Evol. Biosph.* 28, 311–319.

Lin, L.-H., Hall, J., Lippmann-Pipke, J., Ward, J.A., Sherwood Lollar, B., DeFlaun, M., Rothmel, R., Moser, D., Gihring, T.M., Mislowack, B., and Onstott, T.C. (2005) Radiolytic H_2 in continental crust: nuclear power for deep subsurface microbial communities. *Geochem. Geophys. Geosyst.* 6, Q07003.

Marsan, D. (2005) The role of small earthquakes in redistributing crustal elastic stress. *Geophys. J. Int.* 163, 141–151.

Mooney, W.D., Laske, G., and Masters, T.G. (1998) CRUST 5.1: a global crustal model at $5° \times 5°$. *J. Geophys. Res.* 103, 727–747.

Négrel, P. and Casanova, J. (2005) Comparison of the Sr isotopic signatures in brines of the Canadian and Fennoscandian shields. *Appl. Geochem.* 20, 749–766.

Parkes, R.J., Cragg, B.A., and Wellsbury, P. (2000) Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeology J.* 8, 11–28.

Plescia, J.B. (1991) Wrinkle ridges in Lunae Planum Mars: implications for shortening and strain. *Geophys. Res. Lett.* 18, 913–916.

Plescia, J.B. (1993) Wrinkle ridges of Arcadia Planitia, Mars. *J. Geophys. Res.* 98, 15049–15059.

Popov, Y.A., Pevzner, S.L., Pimenov, V.P., and Romushkevich, R.A. (1999) New geothermal data from the Kola superdeep well SG-3. *Tectonophysics* 306, 345–366.

Rickard, D. (1997) Kinetics of pyrite formation by the H_2S oxidation of iron (II) monosulfide in aqueous solutions between 25 and 125°C: the rate equation. *Geochim. Cosmochim. Acta* 61, 115–134.

Rickard, D. and Luther, G.W., III (1997) Kinetics of pyrite formation by the H_2S oxidation of iron (II) monosulfide in aqueous solutions between 25 and 125°C: the mechanism. *Geochim. Cosmochim. Acta* 61, 135–147.

Sherwood Lollar, B., Lacrampe-Couloume, G., Slater, G.F., Ward, J., Moser, D.P., Gihring, T.M., Lin, L.H., and Onstott, T.C. (2006) Unravelling abiogenic and biogenic sources of methane in the Earth's deep subsurface. *Chem. Geol.* 226, 328–339.

Skinner, B.J. (1966) Thermal expansion. In *Handbook of Physical Constants, Geological Society of America, Memoir 97*, Geological Society of America, Boulder, CO, pp. 75–96.

Sleep, N.H. (2006) Mantle plumes from top to bottom. *Earth-Science Reviews* 77, 231–271.

Solomon, S.C. (1987) Secular cooling of the Earth as a source of intraplate stress. *Earth Planet. Sci. Lett.* 83, 153–158.

Stefánsson, A. and Arnórsson, S. (2002) Gas pressures and redox reactions in geothermal fluids in Iceland. *Chem. Geol.* 190, 251–271.

Townend, J. and Zoback, M.D. (2000) How faulting keeps the crust strong. *Geology* 28, 399–402.

Turcotte, D.L. and Schubert, G. (2002) Geodynamics, 2nd ed., Cambridge University Press, Cambridge.

Wachtershauser, G. (1988) Pyrite formation, the first energy source for life: a hypothesis. *System Appl. Microbiol.* 10, 207–210.

Zoback, M.D. and Townend, J. (2001) Implications of hydrostatic pore pressures and high crustal strength for the deformation of intraplate lithosphere. *Tectonophysics* 336, 19–30.

Zoback, M.L., Zoback, M.D., Adams, J., Assumpcao, M., and Bell, S. (1989) Global patterns of tectonic stress. *Nature* 341, 291–298.

Zoback, M.D., Townend, J., and Grollimund, B. (2002) Steady-state failure equilibrium and deformation of intraplate lithosphere. *Int. Geol. Rev.* 44, 383–401.

Address reprint requests to:
Norman H. Sleep
Department of Geophysics
Mitchell Building, Room 366
397 Panama Mall
Stanford University
Stanford, California 94305 USA

E-mail: norm@stanford.edu